REPORT DOCUMENTATION PAG

AFRL-SR-BL-TR-00-

0479

0188

Public reporting burden for this collection of information is estimated to average 1 hour per resp ing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of infor her aspect of this collection of information, including suggestions for reducing this burden, to Washington Headqua s, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budg 503.

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE September 7, 2000	3. REPORT TYPE AND DATES COVERED Final Technical Report, March 1, 1996-Feb. 28,1998

4. TITLE AND SUBTITLE
Final Technical Report, AFOSR F49620-95-1-0214
Statistical Techniques for Modeling, Estimation and Optimization in Distributed Parameter Systems

5. FUNDING NUMBERS
F49620-95-1-0214

6. AUTHOR(S)
Dr. Ben G. Fitzpatrick

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Research in Scientific Computation
Box 8205
North Carolina State University
Raleigh, NC 27695-8205

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Marc Q. Jacobs
AFOSR/NM
801 North Randolph St. Room 732
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
F49620-95-1-0214

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This research program involves mathematical and statistical techniques for using models to analyze experimental data, to aid in experimental design and to optimize system designs. The applications of interest include subsurface contaminant transport, and flow measurement and optimal design in aerospace testing experiments.

20001016 053

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500 DTIC QUALITY INSPECTED 4 Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std. Z39-18

Final Technical Report: F49620-95-1-0214

Statistical Techniques for Modeling, Estimation, and Optimization in Distributed Parameter Systems

Ben G. Fitzpatrick
Center for Research in Scientific Computation
Box 8205
North Carolina State University
Raleigh, NC 27695-8205

Summary. This research program involves mathematical and statistical techniques for using models to analyze experimental data, to aid in experimental design, and to optimize system designs. The applications of interest include subsurface contaminant transport, and flow measurement and optimal design in aerospace testing experiments.

Fate and Transport Research. Research continues on efficient modeling and parameter estimation in heterogeneous media. We have developed new models based on point processes, which allow for "blocky," piecewise constant heterogeneities, and we have tested these methods against tradition kriging estimates as well as our regularized least squares approach. We have also developed transport models based on the statistical characterizations, and we have developed numerical and statistical methods for analyzing unsaturated flow problems for heterogeneous media. Michael Jeffris completed his Ph.D. thesis on simulation of unsaturated flow in heterogeneous media.

Inverse Interferometry. In joint work with Dr. Steve Keeling of AEDC, we have developed and tested numerical algorithms for estimating flowfield density from interferogram data. The algorithms involve a minimum residual cost functional penalized with a total variation regularization term. We have developed techniques for distributing the regularization level in an adaptive manner, to capture better piecewise smooth density profiles.

Turbulence Models and Optical Propagation. The Airborne Laser (ABL) program involves the propagation of laser light over long distances in the atmosphere, and during its travel the light encounters significant turbulence. The effects of the turbulence, which manifests itself through index of refraction variations, are to spread the light beam, thereby reducing its effectiveness. A major challenge for this program is the statistical characterization of the turbulent atmosphere, so that accurate propagation simulations can be made to assess performance of system designs. In collaboration with Dr. Don Washburn of PL/LIA, we have transitioned our statistical methods for site characterization in contaminant transport applications into techniques for understanding the turbulence data obtained by Phillips Lab scientists. These methods allow us to extract regions of stationarity from non-stationary time series data.

Accomplishments.

Subsurface contaminant transport. In this research area, we have made progress in several modeling and parameter estimation problems involving transport in saturated heterogeneous media.

Most of our effort has focused on a type of random process known as a modulated Boolean point process, or MBPP for short, which provides a simple statistical model of "blocky" functions (i.e., functions which are piecewise smooth, with fairly complex discontinuities). These processes are modeled statistically using two probability distributions: one for the number of blocks and the other for the size of the blocks.

A drawback of the standard second-order stationary modeling used for conductivities is that the realizations do not typically "look like the true fields." One expects some small scale variation which resembles a second-order stationary process, with some larger scale variations that are "blockier." A modeling approach that provides such a structure is the modulated point process.

The modeling begins with a (random) number of points in the region of interest, points which represent centers of the "blocks." The next step is the block size and shape. The mathematical formulation of this idea is a function

$$K(x) = \sum_{i=1}^{N} K_i \phi\Big(\frac{x - \xi_i}{r_i}\Big),$$

in which N is a discrete random variable denoting the number of blocks, K_i denotes the "amplitude" of the block, ξ_i denotes the random block centers, r_i denotes the block sizes, and the (deterministic) function ϕ gives the block shape. For example, ϕ can be a Gaussian function, yielding smoothly shaped blocks; alternatively, ϕ can be a box-car function, producing blocks with sharp edges. The function ϕ can also contain anisotropies (imagine a Gaussian function with a non-diagonal covariance in it). This type of process is (when all the K_i values are one) the standard Boolean point process. The modulated process arises from putting a Boolean point process "on top of" the original blocks.

The mathematical procedure for building this model is as follows. For

each center and size above, we have

$$Y_i(x) = k_i \sum_{j=1}^{N_i} \phi_i\Big(\frac{x - \xi_{ij}}{r_{ij}}\Big)$$

with ϕ_i given by

$$\phi_i(s) = \phi\Big(\frac{s - \xi_i}{r_i}\Big).$$

Then the modulated process is given by

$$f(x) = \sum_{i=1}^{N} Y_i(x).$$

In this construction, the new centers ξ_{ij} and radii r_{ij} have a distribution dependent on the values ξ_i and r_i, for each i. Realizations of these processes have the appearance of nonstationarity, but in fact the process is a stationary process. Since it is often mentioned in the literature that field sites do not appear to be stationary, we feel that this modeling tool may be particularly appropriate for hydraulic conductivity modeling. Our results for the NAS site at Columbus AFB are quite encouraging.

We should also remark that, while the realizations of this type of process are dramatically different from standard second-order stationary processes, the covariances can be quite similar. This fact has implications for the 'self-similarity' modeling that is currently being proposed to explain the scale dependence of dispersion parameters.

The estimation problem involves estimating the values of the random quantities N, the block radii and centers, and the intensities. We have developed a Markov Chain Monte Carlo algorithm to generate distributed estimates of the hydraulic conductivity field from observations of the field at discrete points. The method works as follows. One initiates a random walk modeling the center location of the modulator function (i.e., the ξ_i), with a fixed (and large) radius (i.e., the r_i). The random walk is like a simulated annealing search, directed toward regions with datapoints having similar values for the conductivity. Once this point is established, we begin another random walk, and we keep adding modulator sets until we obtain insufficient improvement in the fit. Then we reduce the radius size and begin again.

We continue this procedure until the radii become too small. Of course, "too small" and "insufficient improvement" mean that the user must specify these tolerances. The final stage of this effort is to devise robust and reliable means of assisting the user in specifying these toerances.

One interesting point is that it is a straightforward matter, based on the particular choices of probability distributions for the random quantities in the model, to obtain stationary or non-stationary behavior, as well as the self-similar or 'fractal' behavior proposed by Neumann and others as an explanantion of the scale-effect in dispersion modeling.

In our contaminant transport modeling, major effort has continued on modeling hydrodynamic dispersion. A serious drawback in the application of standard transport models to field scale problems is the determination of the dispersivity parameters. The basic model for a conservative tracer in groundwater is given by

$$nc_t + \nabla \cdot (vc) = \nabla \cdot (D\nabla c),$$

in which $c = c(t, x)$ is the contaminant concentration, n is the porosity, v is the groundwater velocity, and D is the dispersion matrix, given by

$$D = d_0 I + \frac{(a_L - a_T)}{|v|} vv^t + a_T |v| I,$$

in which d_0 is the molecular diffusion coefficient, and a_L and a_T are the longitudinal and transverse dispersivities. The difficulty in applying this model is that the dispersivity parameters appear to depend on the scale of observation: in field scale problems, estimated dispersivity increases as the size of the problem increases.

The approach we have taken to investigate this problem is a stochastic one: modeling the variation in velocity directly and determining a transport model which includes a collection of velocities. It is the variation in velocities at multiple scales which is believed to produce the plume spreading which is (hoped to be) captured by dispersion models.

The basic idea behind the above advection-dispersion equation is that the variation in the velocity field can be modeled as an increase in the (Fickian)

diffusion. Theoretical justifications for such a model rely on large time and small variance asymptotics, which are in general not applicable for many heterogeneous media. Our investigations are along two distinct but related lines. One is related to the point process methods discussed above; the other is a direct Monte Carlo computation.

The point process model begins by imagining the porous medium to be a sequence of well-mixed reactors (cells) connected in series. The modeling assumption that we make is that the time spent in each cell is random. These random times are assumed independent and identically distributed.

We denote by Δx the length of each cell. The time spent in the i^{th} cell is T_i. We denote by μ_t and σ_t^2 the mean and variance of these random times. Since most physical models of porous media treat the velocity (directly, or indirectly, through the hydraulic conductivity) as the "basic" random quantity, we consider random velocities V_i, related to these times by the relationship $\Delta x = V_i T_i$. We denote the velocity mean and variance by μ_v and σ_v^2, respectively. In general, one must know the probability distribution of V_i to determine the mean and variance of T_i, since the relationship between the two is nonlinear. Below we shall give an example of how to relate the time and velocity parameters for lognormally distributed conductivity.

The time required to pass through the first n cells is given by $Z_n = \sum_{i=1}^{n} T_i$. Alternatively, we can determine the cell containing the particle at time t by $M(t) = \inf\{n: Z_n > t\}$. The distance traveled by the particle in this system is then given by $X(t) = \Delta x M(t)$. The probability density of the particle position is the quantity related to the solution of the advection dispersion equation.

The central limit theorem for renewal processes indicates that

$$\frac{M(t) - t/\mu_t}{\sqrt{\sigma_t^2 t/\mu_t^3}} \to N(0,1),$$

where the convergence is convergence in distribution. Of course, to relate this result to the travel distance, $X(t)$, we simply multiply and divide by Δx to obtain

$$\frac{X(t) - \Delta x\, t/\mu_t}{\sqrt{\Delta x^2\, \sigma_t^2 t/\mu_t^3}} \to N(0,1).$$

The asymptotic mean term in this model is given by

$$E(X(t)) \approx \frac{\Delta x\, t}{\mu_t} = \frac{\Delta x\, t}{E(\frac{1}{V_i})},$$

and the asymptotic variance is

$$Var(X(t)) \approx \frac{\Delta x^2\, \sigma_t^2 t}{\mu_t^3}.$$

We point out here that the asymptotic mean contains the term $1/E(1/V_i)$ which is of course different from $E(V_i)$. Our numerical studies indicate that this parameterization compares much better to Monte Carlo simulations that the commonly used approximations based on particle tracking in a second order stationary velocity field.

Another interesting line of inquiry has been in the study of probability models for the velocity that have heavy tails. In fact, using results from the modeling of network traffic, the anomalous dispersion described by Dagan and Neumann can be understood in that context. The basic model of the sequence of mixing regions remains the same, but instead of using only first and second statistical moments, we assume a probability distribution whose tailhas a slow decay rate. In this circumstance, the central limit theorem produces a different limit, whose dispersion coefficient matches field data much more accurately.

As noted above, in order to determine these parameters from the physical parameters (i.e., the hydraulic conductivity's statistical properties), we must know the probability distribution of the velocity field. We have performed these computations using lognormally distributed conductivity processes (which is the most common model in the hydrogeology literature), and we have seen very good agreement with the Monte Carlo simulations (which are discussed below).

Our second approach is to model the plume as being a superposition of many plumes, each following a different velocity field. These different velocity fields are modeled as random fields arising ¿from random field models of the conductivity. The plume is then determined by summing up these

"sub-plumes." The computations involve the Monte Carlo simulation of the hydraulic conductivities, which through Darcy's law generate velocity fields. The advection-diffusion equation (i.e., $D = d_0 I$, no dispersion terms) is solved using these random velocity fields instead of an average velocity, and the "sub-plumes" are superposed to obtain the plume behavior. We have proved convergence theorems for these numerical approximations to capture the mean behavior of the plume, and we have conducted extensive numerical simulations to compare with simpler approximate models found in the literature. Much of this work can be found in our recently submitted paper (in the publications list below). An advantage to this approach (which we are currently investigating) is that one can use a wide variety of random process models for the conductivity. We are in the process of comparing the MBPP and second-order stationary models, from the point of view of transport.

Inverse Interferometry. In this research effort, a collaboration with Dr. Steve Keeling of Sverdrup Technology, Inc., Computation and Modeling Branch, AEDC, Arnold Air Force Base, we have examined some theoretical issues associated with the use of total variation based image reconstruction. Our investigations are motivated by problems of inverse interferometry, in which laser light phase shifts are used to reconstruct medium density profiles in flow field sensing. The reconstruction problem is posed as a residual minimization with total variation regularization applied to handle the inherent ill-posedness. We consider numerical approximations of these penalized minimal residual problems, and analyze some approximation strategies and their properties. The standard definition of total variation leads to inconsistent approximations, with piecewise constant basis functions, so we consider alternative definitions, which preserve the needed compactness and produce convergent approximations.

The general problem to be considered involves a theoretical image $f \in X$, an observed image $g \in Z$, and an observation operator $\Phi : X \to Z$. For instance, $X = L^1(F)$, where F is a bounded flow-field domain in $\mathbf{R}^d$, and $Z = L^1(P)$, where P is a bounded measurement domain in $\mathbf{R}^2$. Alternatively, Z may be a finite dimensional space of *pixel values.* A typical problem involves solving for f in $\Phi f = g$ with observed data $g \in Z$. The problem is treated by minimizing some measure of the residual, $\Phi f - g$, subject to

smoothing constraints which *de-noise* the image, f. In particular, we have studied theoretically and computationally penalized cost functionals of the form

$$J_\mu(f) = \|\Phi f - g\|_{L^1} + \mu TV_1(f),$$

in which f represents the difference between the "far field" fluid density and the density in the region of interest, g represents the interferogram, and the forward operator, Φ, is the interferometry operator

$$\Phi(f)(x,y) = G \int_y^R \frac{f(x,r)}{\sqrt{r^2 - y^2}} r dr,$$

with G being a known constant. This operator is derived assuming axisymmetric flow.

We remark here that the use of the L^1 residual criterion arises for several reasons. One is that L^1 is a natural space for flow-field densities, and the operator Φ lies in $\mathcal{L}(L^1(F), L^1(P))$. Another is that the effectiveness of smoothing via the total variation requires BV to be compactly imbedded in X, a condition met by $X = L^1$ but not $X = L^2$ in dimension $d \geq 2$. Finally, the L^1 criterion is typically more robust in a statistical sense than the L^2 criterion.

The notation TV_1 indicates that our definition of the total variation is not the traditional one. The details of the definition are somewhat tedious; it is most easily explained for smooth functions, for which $TV_1(f) = \int_F |\nabla f(x)|_1 \, dx$ where $|\cdot|_1$ denotes the 1-norm in $\mathbf{R}^2$. The use of this definition of total variation is needed in order to devise consistent numerical approximations for piecewise constant functions. In order to make the cost functional differentiable, we consider the approximation

$$U(f) = \int_F \sqrt{\gamma^2 + |\nabla f|^2} \, dx$$

and in our paper we prove that minimizers of the cost penalized with U converge to those with TV_1 as the penalty, as $\gamma \to 0$.

The focus during this reporting period has been on advancing the numerical work and improving the speed and accuracy of the code. One particular area of effort has been in distributed regularization. Distributed regularization is a technique which uses a spatially varying regularization parameter

to weight the image differently in regions of differing smoothness. We are currently investigating the use of a penalty term of the form

$$\hat{U}(f) = \int_F \frac{\sqrt{\gamma^2 + |\nabla f|^2}}{\sqrt{1 + \omega^2 |\nabla f_0|^2}} \, dx$$

in which f_0 is a "lagged" estimate of the density, and ω is a weight. The function $\mu/\sqrt{1 + \omega^2 |\nabla f_0|^2}$ plays the role of the regularization parameter, becoming nearly zero in regions of large gradients, but remaining near μ when the gradients are small. Numerical experiments with this form of regularization have been very promising.

Turbulence Modeling for Optical Propagation. This project concerns the statistical characterization of atmospheric turbulence, from the perspective of optical propagation through the atmosphere. We are currently collaborating with PL/LIA scientists on several topics, including the estimation of mean and variance parameters for temperature fluctuations in the atmosphere, determination of probability densities for these parameters, developing confidence bounds on parameter estimates, and finding efficient simulation methods.

The point process models we have developed for groundwater applications appear to be applicable in atmospheric turbulence, in that observed data often exhibit behavior that appears to be nonstationary in the mean. Our MBPP models ppear to fit the data quite well, giving us a reasonable decomposition of the time series into regions of stationarity.

Of crucial importance in the analysis of laser propagation is the accurate simulation of atmospheric variability. In order to perform these simulations, we need not only estimates of parameters in the statistical model of turbulence, but also an accurate sampling distribution. We are applying nonparametric density estimation methods developed for groundwater studies to the problem of the sampling distribution for the phase screens (index of refraction characterizations) simulated from turbulent atmosphere statistics.

Personnel supported. Assoicated with this research effort are several coworkers. Dr. Christian Wypasek, a postdoctoral researcher, has collaborated in the statistical modeling of hydraulic conductivity. Two graduate

students, Jeff Butera (supported by a DOE Computational Sciences Fellowship) and Mike Jeffris (both supported by Dept. of Education GAANN Fellowships) have participated in various aspects of parameter estimation in groundwater flow problems. Butera has recently accepted a tenure track appointment at High Point University, while Wypasek has been employed since 1 July 97 at GE Financial Services. Mike Horton, a graduate student in mathematics supported on this grant, is currently investigating on certain statistical aspects of the ABL work. Anand and Aravind Ramachandran, two local high school students studying mathematics at NCSU, are also programming some smoothing algorithms for use on the ABL dataset.

Publications. We have the following publications to report:

- "Estimation of Time Dependent Parameters in General Parabolic Evolution Systems," by A.S. Ackleh and Ben G. Fitzpatrick, *J. Math. Anal. Appl.* **203**, 1996, pp. 464-480.
- "Estimation of Discontinuous Parameters in General Nonautonomous Parabolic Systems," by by A.S. Ackleh and Ben G. Fitzpatrick, *Kybernetica,* **32**, number 6, 1996, pp 543-556.
- "An Adaptive Change Detection Scheme for a Nonlinear Beam Model," by M. A. Demetriou and B. G. Fitzpatrick, *Kybernetica,* **33**, number 1, 1997, pp 103-120.
- "On Approximation in Total Variation Penalization for Image Reconstruction and Inverse Problems," by Ben G. Fitzpatrick and Steven L. Keeling, accepted, *J. Num. Func. Anal. Opt.*
- "Estimation of Groundwater Flow Parameters Using Least Squares," by Kendall R. Bailey and Ben G. Fitzpatrick, accepted, *Math. Comp. Modeling.*
- "Modeling Aggregation and Growth Processes in an Algal Population: Analysis and Computations" by A. S. Ackleh and B. G. Fitzpatrick, *J. Math. Bio.* **35**, 1997, pp. 480-502. **Interactions/Transitions.**
- "On Continuous Dependence under Approximation for Groundwater Flow Models with Distributed and Pointwise Observations," by Ben G.

Fitzpatrick and Michael A. Jeffris, *Discrete and Continuous Dynamical Systems,* **2**, no. 1, 1996, pp. 141-149.

- "Sampling Distribution of Approximate Errors for Least Squares Identification," by G. Yin, Ben G. Fitzpatrick, and K. Yin, accepted, *Stochastic Anal. Appl.*

- "Dispersion Modeling and Simulation in Subsurface Contaminant Transport," by Jeffrey V. Butera, Ben G. Fitzpatrick, and Christian J. Wypasek, submitted to *Math. Models and Methods in Appl. Sci.*

a. Meetings and seminars. The Principal Investigator has given two talks at Phillips Lab worshops on Atmospheric Turbulence and Optical Propagation (Dec 96 and July 97) and a talk at the 1997 AFOSR/ARO Environmental Quality contractors' meeting. He attended the Electricity and Magnetism AFOSR workshop at Brooks AFB (Jan 97). He has given colloquium talks at Worcester Polytechnic Institute (Mar 97) and at the University of Southern Louisiana (Apr 97), and an invited presentation at the 1997 Barrett Lectures, University of Tennessee (Mar 97). He gave a seminar talk at D. H. Wagner, Associates, on the inverse interferometry effort (July 97). Dr. Christian Wypasek and Mr. Jeff Butera gave talks at the SIAM Conference on Geosciences (June 1997).

b. Consultative functions. The Principal Investigator visited Kirtland AFB several times during this report period (Oct, Dec 96; April, May, June, July 97) to discuss various aspects of the data analysis and modeling efforts for the ABL work. In January of 1997, he visited Brooks AFB to continue collaborations with scientists on E& M control problems. Several trips were made to Arnold Engineering Development Center (Nov 96; May, July, Aug 97) to continue collaborations on the inverse interferometry and design optimization efforts.